Extrapolation of stable random fields[☆]

Wolfgang Karcher[a], Elena Shmileva[b], Evgeny Spodarev[a]

[a]*Ulm University, Institute of Stochastics, Helmholtzstr. 18, 89081 Ulm, Germany*
[b]*St. Petersburg State University, Chebyshev Laboratory, 14th Line 29b, St. Petersburg 199178, Russia*

Abstract

In this paper, we discuss three extrapolation methods for α-stable random fields with $\alpha \in (1, 2]$. We justify them, giving proofs of the existence and uniqueness of the solutions for each method and providing sufficient conditions for path continuity. Two methods are based on minimizing the variability of the difference between the predictor and the theoretical value, whereas in the third approach we provide a new method that maximizes the covariation between these two quantities.

Keywords: Extrapolation, random field, stable distribution
2000 MSC: 60G60, 60G25, 62M20

1. Introduction

In many applications, Gaussian random fields are chosen as a model for the regionalized variables. However, natural phenomena may exhibit heavy tails such that the assumption of Gaussianity is not reasonable any more. In this case, stable random fields may be a more appropriate choice as they still possess the desirable properties of Gaussian random fields, but allow for heavy tails.

In this paper, we are dealing with the problem of extrapolating an α-stable random field. Namely, based on a given set of data points $\{x(t_i)\}_{i=1}^{n}$ at the locations $t_1, \ldots, t_n, n \in \mathbb{N}$, which are part of a realization of an α-stable random field, we estimate the unknown value $x(t_0)$ at the location $t_0 \notin \{t_1, \ldots, t_n\}$. We present three linear predictors of the form

$$\widehat{x(t_0)} = \sum_{i=1}^{n} \lambda_i x(t_i), \quad t_0 \in \mathbb{R}^d. \tag{1}$$

The weights $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ in (1) will be chosen in such a way that the predictor $\widehat{x(t_0)}$ has certain desirable properties. Two predictors are based on ideas of previous work in the literature, namely we generalize the existing methods to a wide range of α-stable random fields and investigate the properties of existence, uniqueness, exactness, unbiasedness and continuity of the predictors in detail. We also propose a third new extrapolation method which maximizes the covariation dependence measure between the estimated and theoretical value.

A 2-stable random field is a Gaussian random field. A widely used extrapolation technique for Gaussian random fields is kriging which was invented by Krige in 1951, see [11]. The idea behind kriging is to find a linear predictor which minimizes the L^2-distance between the predictor and the theoretical value. We refer to [19] for further details about kriging. Since the second moment of an α-stable random variable does not exist when $\alpha < 2$, the kriging technique cannot be applied to α-stable random fields so that one has to look for alternative extrapolation methods.

There is a series of papers in the literature dealing with the extrapolation problem of particular stable random processes and fields. In [16], maximum likelihood estimation and conditional simulation of the fractional stable

motion is used to determine the optimal value at an unobserved location. Linear prediction of discrete α-stable processes based on minimizing the dispersion between the predictor and the theoretical value is discussed in [1, 2, 3, 7, 8, 9]. Finally, a modified dispersion measure is minimized in [13] in order to obtain a linear prediction of discrete α-stable processes with integral representation.

The paper is organized as follows. In Section 2, we provide some basics about α-stable random fields. In Section 3, we present the three extrapolation methods for $\alpha \in (1, 2]$ (least scale linear predictor, covariation orthogonal linear predictor and maximum of covariation predictor) and analyze their properties. The proofs of the results can be found in the Appendix.

2. Preliminaries

In this section, we provide some basic definitions and results for α-stable random variables, random vectors and random fields. For a detailed introduction to α-stable distributions and processes, we refer to [18].

Definition 1 (α-stable random variable). Let $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$, and $\mu \in \mathbb{R}$. A random variable X is said to have an α-*stable distribution* with parameters σ, β, and μ, if its characteristic function φ_X is given by

$$\varphi_X(\theta) = \begin{cases} \exp\left\{-\sigma^\alpha |\theta|^\alpha \left(1 - i\beta(\text{sign}(\theta)) \tan\frac{\pi\alpha}{2}\right) + i\mu\theta\right\}, & \alpha \neq 1, \\ \exp\left\{-\sigma|\theta| \left(1 + i\beta\frac{2}{\pi}(\text{sign}(\theta)) \ln|\theta|\right) + i\mu\theta\right\}, & \alpha = 1. \end{cases}$$

We briefly write $X \sim S_\alpha(\sigma, \beta, \mu)$. The parameters σ, β and μ determine the scale, the skewness and the shift of the random variable, respectively, whereas the parameter α specifies the tail behavior of the distribution of X. We notice that $S_2(\sigma, 0, \mu) = \mathcal{N}(\mu, 2\sigma^2)$, where $\mathcal{N}(\mu, 2\sigma^2)$ denotes the normal distribution with mean μ and variance $2\sigma^2$.

If $X \sim S_\alpha(\sigma, \beta, \mu)$ with $\alpha \in (0, 2)$, then

$$\mathbb{E}|X|^p < \infty \quad \text{for any} \quad p \in (0, \alpha), \tag{2}$$

$$\mathbb{E}|X|^p = \infty \quad \text{for any} \quad p \in [\alpha, \infty). \tag{3}$$

Moreover if $\alpha \in (0, 2)$ and $\beta = 0$ in the case $\alpha = 1$, then for every $p \in (0, \alpha)$ there exists a constant $c_{\alpha,\beta}(p) > 0$ such that

$$(\mathbb{E}|X|^p)^{1/p} = c_{\alpha,\beta}(p)\sigma. \tag{4}$$

An α-stable random vector is defined as follows. Let $d \in \mathbb{N}$ and $\langle \cdot, \cdot \rangle$ denote the Euclidean scalar product in $\mathbb{R}^d$.

Definition 2 (α-stable random vector). Let $0 < \alpha < 2$. We say that a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ in $\mathbb{R}^d$ has a *(multivariate) α-stable distribution* if there exists a finite measure Γ on the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ and a vector $\boldsymbol{\mu}$ in $\mathbb{R}^d$ such that the characteristic function $\varphi_{\boldsymbol{X}}$ is given by

$$\varphi_{\boldsymbol{X}}(\boldsymbol{\theta}) = \begin{cases} \exp\left\{-\int_{\mathbb{S}^{d-1}} |\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle|^\alpha \left(1 - i\text{sign}(\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle) \tan\frac{\pi\alpha}{2}\right) \Gamma(d\boldsymbol{s}) + i\langle \boldsymbol{\theta}, \boldsymbol{\mu} \rangle\right\}, & \alpha \neq 1, \\ \exp\left\{-\int_{\mathbb{S}^{d-1}} |\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle| \left(1 + i\frac{2}{\pi}\text{sign}(\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle) \ln|\langle \boldsymbol{\theta}, \boldsymbol{s} \rangle|\right) \Gamma(d\boldsymbol{s}) + i\langle \boldsymbol{\theta}, \boldsymbol{\mu} \rangle\right\}, & \alpha = 1. \end{cases}$$

The pair $(\Gamma, \boldsymbol{\mu})$ is unique. The *spectral measure* Γ on the unit sphere $\mathbb{S}^{d-1}$ contains most information about $\boldsymbol{X}$ such as the dependence structure and distributional properties of its components. The vector $\boldsymbol{\mu}$ determines the shift with respect to the origin. We say that Γ is *concentrated on a great sub-sphere* of $\mathbb{S}^{d-1}$ if it is concentrated on the intersection of $\mathbb{S}^{d-1}$ with a $(d-1)$-dimensional linear subspace. If Γ is not concentrated on a great sub-sphere of $\mathbb{S}^{d-1}$, then $\boldsymbol{X}$ is called *full-dimensional*. Otherwise, $\boldsymbol{X}$ is called *singular*.

Lemma 1 (Linear dependence). *Let $d \in \mathbb{N}$ and consider a d-dimensional α-stable random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$. Let $\boldsymbol{\mu} = (0, \ldots, 0)^\top \in \mathbb{R}^d$. $\boldsymbol{X}$ is singular if and only if $\sum_{i=1}^d c_i X_i = 0$ almost surely for some $(c_1, \ldots, c_d)^\top \in \mathbb{R}^d \setminus \{0\}$.*

Definition 3 (Linear regression). Let $n \in \mathbb{N}$ and $(X_1, \ldots, X_n, X_{n+1})^\top$ be an α-stable random vector. The regression of X_{n+1} on $(X_1, \ldots, X_n)^\top$ is called *linear* if there exist some constants $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$\mathbb{E}(X_{n+1}|X_1, \ldots, X_n) = \sum_{i=1}^n c_i X_i \quad a.s. \tag{5}$$

2

Let sp(X) denote the linear span of the α-stable random vector X.

Definition 4 (Multiple regression property). Let $n \in \mathbb{N}$. The vector X has the *multiple regression property* if for all random variables $Y_1, \ldots, Y_{n+1} \in \text{sp}(X)$

$$\mathbb{E}(Y_{n+1}|Y_1, \ldots, Y_n) \in \text{sp}\left((Y_1, \ldots, Y_n)^\mathsf{T}\right).$$

A random vector X in $\mathbb{R}^d$ is called *symmetric* if $\mathbb{P}(X \in A) = \mathbb{P}(-X \in A)$ for any Borel set $A \in \mathbb{R}^d$. If X is a symmetric α-stable random vector in $\mathbb{R}^d$, its characteristic function φ_X is given by

$$\varphi_X(\boldsymbol{\theta}) = \exp\left\{-\int_{\mathbb{S}^{d-1}} |\langle \boldsymbol{\theta}, s\rangle|^\alpha \Gamma(ds)\right\}, \quad \boldsymbol{\theta} \in \mathbb{R}^d.$$

For symmetric α-stable distributions, we use the standard abbreviation $S\alpha S$.

The dependence of two α-stable random variables cannot be studied by using the covariance because of the absence of the second moments if $\alpha < 2$. Thus, we have to use a different dependence measure. The most popular dependence measure between two α-stable random variables is the covariation. It is only defined for α-stable random variables with $1 < \alpha \leq 2$. Let $a^{<p>} := |a|^p \text{sign}(a)$ denote the signed power for $a \in \mathbb{R}$ and $p \in \mathbb{R}$.

Definition 5 (Covariation). Let $\alpha \in (1, 2]$ and Γ be the spectral measure of the α-stable random vector $X = (X_1, X_2)^\mathsf{T}$. The *covariation* of X_1 on X_2 is the real number

$$[X_1, X_2]_\alpha := \int_{\mathbb{S}^1} s_1 s_2^{<\alpha-1>} \Gamma(ds_1, ds_2).$$

The covariation is linear in the first entry, but in general not in the second. It is clear that $[X_1, X_2]_\alpha = 0$ if X_1 and X_2 are independent. For $\alpha = 2$, we have $[X_1, X_2]_2 = 1/2 \cdot \text{Cov}(X_1, X_2)$, where $\text{Cov}(X_1, X_2)$ denotes the covariance between X_1 and X_2, see [18, Example 2.7.2].

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space and $L_0(\Omega)$ be the set of all real random variables defined on it. Furthermore, let $(E, \mathcal{E}, m)$ be an arbitrary measurable space, $\beta : E \to [-1, 1]$ be a measurable function and $\mathcal{E}_0 := \{A \in \mathcal{E} : m(A) < \infty\}$.

Definition 6 (α-stable random measure). An independently scattered σ-additive set function $M : \mathcal{E}_0 \to L_0(\Omega)$ such that for each $A \in \mathcal{E}_0$

$$M(A) \sim S_\alpha\left((m(A))^{1/\alpha}, \frac{\int_A \beta(x)m(dx)}{m(A)}, 0\right)$$

is called α-*stable random measure* on $(E, \mathcal{E})$ with control measure m and skewness β.

Independent scatteredness means that for any collection $A_1, A_2, \ldots, A_n$, $n \in \mathbb{N}$, of disjoint sets belonging to $\mathcal{E}_0$, the random variables $M(A_1), \ldots, M(A_n)$ are independent, whereas σ-*additivity* means that if $\bigcup_{j=1}^\infty A_j \in \mathcal{E}_0$ for a sequence of disjoint sets $A_1, A_2, \ldots \in \mathcal{E}_0$, then $M(\bigcup_{j=1}^\infty A_j) = \sum_{j=1}^\infty M(A_j)$ almost surely.

A collection $X = \{X(t), t \in \mathbb{R}^d\}$ of α-stable random variables is called α-*stable random field*. We consider random fields of the form

$$X(t) = \int_E f_t(x)M(dx), \quad t \in \mathbb{R}^d, \quad d \in \mathbb{N}, \tag{6}$$

where $f_t : E \to \mathbb{R}$ are measurable functions such that $\int_E |f_t(x)|^\alpha m(dx) < \infty$ and in the case $\alpha = 1$ additionally $\int_E |f(x)\beta(x) \ln |f(x)| |m(dx) < \infty$. The stochastic integral in (6) is defined in the natural way by approximating the functions f_t by simple functions. We notice that the random field X in (6) is α-stable because its finite-dimensional distributions are α-stable. In particular, the scale parameter $\sigma_{X(t)}$ of $X(t)$ is given by

$$\sigma_{X(t)}^\alpha = \int_E |f_t(x)|^\alpha m(dx). \tag{7}$$

If $\beta(x) = 0$ for all $x \in E$, then the corresponding α-stable random measure has $S\alpha S$ values and is called $S\alpha S$ *measure*. In this case, the finite-dimensional distributions of X are symmetric α-stable, so we have that X is an $S\alpha S$ random field.

Random fields with integral representation (6) comprise a rich subclass of α-stable random fields, namely random fields which are separable in probability, see [18, Chapter 13]. In particular, all stochastically continuous α-stable random fields are separable in probability.

Let $X_j = \int_E f_j(x)M(dx)$, $j = 1, 2, \ldots$, and $X = \int_E f(x)M(dx)$. Then for $\alpha \neq 1$, it can be shown that

$$\operatorname*{plim}_{j \to \infty} X_j = X \quad \Leftrightarrow \quad \lim_{j \to \infty} \sigma_{X_j - X} = 0, \tag{8}$$

where plim denotes convergence in probability, cf. [18, Proposition 3.5.1].

In [4, Theorem 1 and Remark 1], it is shown that for $t_0, t_1, \ldots, t_n \in \mathbb{R}^d$ it holds

$$\mathbb{E}(X(t_0)|X(t_1), \ldots, X(t_n)) = \sum_{i=1}^{n} c_i X(t_i)$$

if and only if for all $t \in \mathbb{R}^n$

$$\int_{\mathbb{S}^n} (v - \langle c, u \rangle) \langle t, u \rangle^{<\alpha-1>} \Gamma(dv, du) = 0, \tag{9}$$

where Γ is the spectral measure of the random vector $(X(t_0), X(t_1), \ldots, X(t_n))^\mathsf{T}$. If X is a moving average process, another criterion based on kernel functions for the regression of $X(t_0)$ on $(X(t_1), \ldots, X(t_n))^\mathsf{T}$ to be linear is given in [4, Section 4.4.].

Lemma 2. *Consider a d-dimensional α-stable random vector $X = (X_1, \ldots, X_d)^\mathsf{T}$ with integral representation*

$$X = \left(\int_E f_1(x)M(dx), \ldots, \int_E f_d(x)M(dx) \right)^\mathsf{T}.$$

Then X is singular if and only if $\sum_{i=1}^{d} c_i f_i(x) = 0$ m-almost everywhere for some vector $(c_1, \ldots, c_d)^\mathsf{T} \in \mathbb{R}^d \setminus \{0\}$.

Lemma 2 follows from Lemma 1 and the fact that $\sigma^\alpha_{\sum_{i=1}^{d} c_i X_i} = \int_E \left| \sum_{i=1}^{d} c_i f_i(x) \right|^\alpha m(dx)$.

Definition 7 (Covariation function). Let $1 < \alpha \leq 2$ and X be an α-stable random field. The function $\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$\kappa(s, t) = [X(s), X(t)]_\alpha, \quad s, t \in \mathbb{R}^d,$$

is called the *covariation function* of X.

For any α-stable random field X of the form (6) with $1 < \alpha \leq 2$, its covariation function is given by

$$\kappa(s, t) = \int_E f_s(x) f_t(x)^{<\alpha-1>} m(dx). \tag{10}$$

A random vector $(X, Y)^\mathsf{T}$ is called *associated* if, for any functions $f, g : \mathbb{R}^2 \to \mathbb{R}$ which are non-decreasing in each argument, one has

$$\operatorname{Cov}(f(X, Y), g(X, Y)) \geq 0 \tag{11}$$

whenever the covariance exists. It is called *negatively associated* if for any functions $f, g : \mathbb{R} \to \mathbb{R}$ which are non-decreasing, one has

$$\operatorname{Cov}(f(X), g(Y)) \leq 0$$

whenever the covariance exists. Notice that the condition $\operatorname{Cov}(f(X, Y), g(X, Y)) \leq 0$ does not make sense because $\operatorname{Cov}(g(X, Y), g(X, Y)) \geq 0$ for any non-degenerate random vector $(X, Y)^\mathsf{T}$.

Lemma 3 (Association, kernel functions)**.** *Let $0 < \alpha < 2$ and $(X, Y)^\mathsf{T}$ be an α-stable random vector with integral representation $(X, Y)^\mathsf{T} = \left(\int_E f_1(x)M(dx), \int_E f_2(x)M(dx)\right)^\mathsf{T}$. Then $(X, Y)^\mathsf{T}$ is associated (negatively associated) if and only if $f_1 f_2 \geq 0$ ($f_1 f_2 \leq 0$) m-almost everywhere.*

Corollary 4 (Decomposition of a stable random vector)**.** *Let $0 < \alpha < 2$ and $(X, Y)^\mathsf{T}$ be an α-stable random vector. Then there exist α-stable random variables X_1, X_2, Y_1, Y_2 such that*

$$X = X_1 + X_2 \quad and \quad Y = Y_1 + Y_2 \quad a.s.,$$

$(X_1, Y_1)^\mathsf{T}$ is associated, $(X_2, Y_2)^\mathsf{T}$ is negatively associated and the components of each of the random vectors $(X_1, X_2)^\mathsf{T}$, $(X_1, Y_2)^\mathsf{T}$, $(Y_1, X_2)^\mathsf{T}$, $(Y_1, Y_2)^\mathsf{T}$ are independent.

Due to Corollary 4, the dependence relation between two α-stable random variables can be reduced to the association and negative association relations between their parts. The following corollary provides a justification for using the covariation as a dependence measure. It is a direct consequence of Lemma 3 and formula (10).

Corollary 5 (Association, covariation)**.** *Let $1 < \alpha < 2$ and $(X, Y)^\mathsf{T}$ be an α-stable random vector. If $(X, Y)^\mathsf{T}$ is associated (negatively associated), then $[X, Y]_\alpha \geq 0$ ($[X, Y]_\alpha \leq 0$).*

The covariation is connected to the mixed moments by the following lemma.

Lemma 6. *Let $1 < \alpha < 2$ and suppose that $(X, Y)^\mathsf{T}$ is an α-stable random vector with spectral measure Γ such that $X \sim S_\alpha(\sigma_X, \beta_X, 0)$ and $Y \sim S_\alpha(\sigma_Y, \beta_Y, 0)$. For $1 \leq p < \alpha$, it holds*

$$\frac{\mathbb{E}\left(XY^{<p-1>}\right)}{\mathbb{E}|Y|^p} = \frac{[X, Y]_\alpha(1 - c \cdot \beta_Y) + c \cdot (X, Y)_\alpha}{\sigma_Y^\alpha}, \tag{12}$$

where $(X, Y)_\alpha := \int_{\mathbb{S}^1} s_1 |s_2|^{\alpha-1}\Gamma(ds)$ and

$$c := c_{\alpha,p}(\beta_Y) = \frac{\tan(\alpha\pi/2)}{1 + \beta_Y^2 \tan^2(\alpha\pi/2)}\left[\beta_Y \tan(\alpha\pi/2) - \tan\left(\frac{p}{\alpha}\arctan(\beta_Y \tan(\alpha\pi/2))\right)\right].$$

If Y is symmetric, i. e. $\beta_Y = 0$, then $c = 0$.

A random vector $\boldsymbol{X}$ is called *sub-Gaussian* if $\boldsymbol{X} \overset{d}{=} A^{1/2}\boldsymbol{G}$, where $A \sim S_{\alpha/2}((\cos(\pi\alpha/4))^{2/\alpha}, 1, 0)$ and $\boldsymbol{G}$ is a zero mean Gaussian vector independent of A. This multivariate distribution is a particular case of the Gaussian scale mixture and the elliptical distributions. Consider a stationary zero mean Gaussian random field $G = \{G(t), t \in \mathbb{R}^d\}$ with a positive definite covariance function C. Assume that the random variable A is independent of G. The random field

$$X \overset{d}{=} \{A^{1/2}G(t), t \in \mathbb{R}^d\}$$

is called *sub-Gaussian random field*. In [18, Example 2.7.4], it is shown that for sub-Gaussian random fields, the covariation function is given by

$$\kappa(s, t) = 2^{-\alpha/2}C(s - t)C(0)^{(\alpha-2)/2}. \tag{13}$$

3. Extrapolation methods

Let X be an α-stable random field of the form (6) and $1 < \alpha \leq 2$. In Sections 3.1–3.3, we present three approaches that yield the weights $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ of the linear predictor

$$\widehat{X(t_0)} = \sum_{i=1}^n \lambda_i X(t_i) \tag{14}$$

for the value $X(t_0)$ at the location $t_0 \in \mathbb{R}^d$ based on the values $X(t_1), \ldots, X(t_n)$ at the neighboring locations $t_1, \ldots, t_n$, $n \in \mathbb{N}$.

A predictor $\widehat{X(t_0)}$ for $X(t_0)$ is called *exact* if $\widehat{X(t_0)} = X(t_0)$ almost surely whenever $t_0 = t_i$ for some $i \in \{1, \ldots, n\}$. We say that $\widehat{X(t_0)}$ is *unbiased* if $\mathbb{E}(\widehat{X(t_0)} - X(t_0)) = 0$. It follows from representation (6) that $\mathbb{E}X(t) \equiv 0$ for each $t \in \mathbb{R}^d$, hence the predictor (14) is always unbiased.

For each $t_0 \in \mathbb{R}^d$, the linear prediction yields a set of weights $\lambda_1, \ldots, \lambda_n$. Let us consider these weights as functions $\lambda_i : \mathbb{R}^d \to \mathbb{R}$ of the locations $t \in \mathbb{R}^d$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space of the random field. If the functions λ_i, $i = 1, \ldots, n$, are continuous, the extrapolator $\hat{X} : \mathbb{R}^d \times \Omega \to \mathbb{R}$ given by

$$\hat{X}(t, \omega) = \sum_{i=1}^{n} \lambda_i(t) X(t_i, \omega), \quad t \in \mathbb{R}^d,$$

is continuous for each $\omega \in \Omega$. In this case, we say that the predictor is *continuous*.

3.1. Least Scale Linear (LSL) predictor

We noticed that the idea behind kriging is the minimization of the prediction variance. Since the variance of a Gaussian random variable is twice the scale parameter of a 2-stable random variable, an obvious generalization of the kriging technique is the minimization of the prediction scale

$$\sigma^{\alpha}_{\widehat{X(t_0)} - X(t_0)} = \int_E \left| f_{t_0}(x) - \sum_{i=1}^{n} \lambda_i f_{t_i}(x) \right|^{\alpha} m(dx) \to \min \tag{15}$$

with respect to $\lambda_1, \ldots, \lambda_n$. Here, $\sigma_{\widehat{X(t_0)} - X(t_0)}$ is the scale parameter of the α-stable random variable $\widehat{X(t_0)} - X(t_0)$, cf. (7). The predictor $\widehat{X(t_0)}$ based on a solution of this minimization problem is called *Least Scale Linear (LSL) predictor*.

The LSL predictor can be further justified as follows. Let $\{\widehat{X_m(t_0)}\}_{m \in \mathbb{N}}$ be a sequence of predictors for $X(t_0)$. First by (8), $\widehat{X_m(t_0)}$ tends to $X(t_0)$ in probability as $m \to \infty$ if and only if $\sigma_{\widehat{X_m(t_0)} - X(t_0)}$ tends to zero as $m \to \infty$. Second by (4), the mean p-error between $\widehat{X_m(t_0)}$ and $X(t_0)$ tends to zero for every $p \in (0, \alpha)$ as $m \to \infty$ if and only if $\sigma_{\widehat{X_m(t_0)} - X(t_0)}$ tends to zero as $m \to \infty$. That is why it is reasonable to minimize $\sigma^{\alpha}_{\widehat{X(t_0)} - X(t_0)}$ in (15).

We now derive the necessary conditions for the existence of the LSL predictor.

Lemma 7. *Let M be an α-stable random measure. Then*

$$\frac{\partial \sigma^{\alpha}_{\widehat{X(t_0)} - X(t_0)}}{\partial \lambda_j} = \alpha \left[X(t_j), \sum_{i=1}^{n} \lambda_i X(t_i) - X(t_0) \right]_{\alpha}, \quad j = 1, \ldots, n.$$

Thus, a solution of the minimization problem (15) resolves the following system of equations

$$\left[X(t_j), X(t_0) - \sum_{i=1}^{n} \lambda_i X(t_i) \right]_{\alpha} = 0, \quad j = 1, \ldots n, \tag{16}$$

which can be written as

$$\int_E f_{t_j}(x) \left(f_{t_0}(x) - \sum_{i=1}^{n} \lambda_i f_{t_i}(x) \right)^{<\alpha - 1>} m(dx) = 0, \quad j = 1, \ldots n, \tag{17}$$

by (10) and the linearity of α-stable integrals. Notice that (17) is nonlinear in $\lambda_1, \ldots, \lambda_n$ if $\alpha < 2$. Therefore, numerical methods have to be applied in this case in order to solve this minimization problem, for example the Levenberg-Marquardt algorithm, see [15].

3.1.1. Properties of the LSL predictor

It is clear that the set of all LSL predictors is closed and convex, see also [6, p. 59].

Theorem 8. *The LSL predictor has the following properties:*

6

1. *The LSL predictor exists, is unique and exact.*

2. *If the random field X is stochastically continuous and the random vector $(X(t_1), \ldots, X(t_n))^\top$ is full-dimensional, then the LSL predictor is continuous.*

Remark 1. The first and second assertion of Theorem 8 follow from the properties of the best approximation in $L^\alpha(E, m)$-spaces for $1 \le \alpha \le 2$. Notice that if $0 < \alpha < 1$, the quantity $\|f\|_{L^\alpha(E,m)} = (\int_E |f(x)|^\alpha m(dx))^{1/\alpha}$ is not a norm anymore (there is neither the triangle inequality nor Minkowski's inequality). Thus for $0 < \alpha < 1$, the existence and uniqueness of the LSL estimator requires additional investigation.

3.2. Covariation Orthogonal Linear (COL) predictor

Let X be an α-stable random field. We have seen that the LSL predictor can be only calculated by numerical methods because the system of equations (17) is nonlinear. The linearity of the covariation in the first entry suggests to exchange both entries in (16) such that we get the system of linear equations

$$\left[X(t_0), X(t_j)\right]_\alpha - \sum_{i=1}^n \lambda_i \left[X(t_i), X(t_j)\right]_\alpha = 0, \quad j = 1, \ldots, n. \tag{18}$$

Since the covariation of $X(t_0) - \sum_{i=1}^n \lambda_i X(t_i)$ on $X(t_j)$ is zero for $j = 1, \ldots, n$, the linear predictor $\widehat{X(t_0)} = \sum_{i=1}^n \lambda_i X(t_i)$ based on any solution $(\lambda_1, \ldots, \lambda_n)$ of (18) is called after [8] *Covariation Orthogonal Linear (COL) predictor*.

The unknown quantities $[X(t_0), X(t_j)]_\alpha$ and $[X(t_i), X(t_j)]_\alpha$ in (18) can be estimated either by using (12) in the symmetric case or by estimating the kernel functions and using (10) in the non-symmetric case. By (18), the COL predictor is the solution of the following system of equations

$$[X(t_0), X(t_i)]_\alpha = [\widehat{X(t_0)}, X(t_i)]_\alpha, \quad i = 1, \ldots, n,$$

which is equivalent to

$$\mathbb{E}\left(X(t_0) X(t_i)^{<p-1>}\right) = \mathbb{E}\left(\widehat{X(t_0)} X(t_i)^{<p-1>}\right), \quad i = 1, \ldots, n,$$

in the case of an $S\alpha S$ random field by (12).

If the similarity of two α-stable random variables is measured by the covariation, the first system of equations ensures that the similarity of the theoretical value $X(t_0)$ and $X(t_i)$ is exactly equal to the similarity of the predictor $\widehat{X(t_0)}$ and $X(t_i)$. The second system of equations guarantees that the "mixed" moments between the theoretical value $X(t_0)$ and $X(t_i)$ agree with the "mixed" moments between the predictor $\widehat{X(t_0)}$ and $X(t_i)$.

For a detailed treatment of the COL predictor of discrete stationary ARMA α-stable processes, we refer to [8]. The following lemma connects linear regression with the COL predictor, see [18, Corollary 4.1.3].

Lemma 9. *If the regression of $X(t_0)$ on the random vector $(X(t_1), \ldots, X(t_n))^\top$ is linear, i. e. there exists some $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ such that $\mathbb{E}(X(t_0)|X(t_1), \ldots, X(t_n)) = \sum_{i=1}^n \lambda_i X(t_i)$ almost surely, then the vector $(\lambda_1, \ldots, \lambda_n)^\top$ is a solution of the system of equations (18).*

The regression of $X(t_0)$ on $X(t_1)$ is always linear, see [18, Theorem 4.1.2]. If $n \ge 2$, then $(X(t_0), X(t_1), \ldots, X(t_n))^\top$ has the multiple regression property if and only if $(X(t_0), X(t_1), \ldots, X(t_n))^\top$ is Gaussian or sub-Gaussian, see [18, Proposition 4.1.7]. Thus for $n \ge 2$, Lemma 9 can be applied to Gaussian and sub-Gaussian random fields.

In the next subsections, we analyze existence, uniqueness and continuity of the COL predictor and concentrate on moving average, Gaussian and sub-Gaussian random fields. Only for these kinds of α-stable random fields, we know how to derive reasonable sufficient conditions such that the properties hold. Notice that for $0 < \alpha < 2$, any stationary $S\alpha S$ random field X can be decomposed into three independent parts $X = X^{(1)} + X^{(2)} + X^{(3)}$, where $X^{(1)}$ is a mixed moving average, $X^{(2)}$ is harmonizable, and $X^{(3)}$ is a stationary $S\alpha S$ process without mixed moving average or harmonizable components, see [17]. Moving averages are a subclass of mixed moving averages, while sub-Gaussian random fields belong to the classes $X^{(2)}$ or $X^{(3)}$, cf. [18, Theorem 4.7.5 and Theorem 6.6.5].

Exactness of the COL predictor holds for all kinds of α-stable random fields which directly follows from the corresponding system of equations (18).

3.2.1. Properties of the COL predictor for Gaussian and sub-Gaussian random fields

Let us first consider a stationary zero mean Gaussian random field G with a positive definite covariance function C. The system of linear equations (18) becomes

$$
\begin{pmatrix} C(0) & \cdots & C(t_n - t_1) \\ \vdots & \ddots & \vdots \\ C(t_n - t_1) & \cdots & C(0) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(t_0 - t_1) \\ \vdots \\ C(t_0 - t_n) \end{pmatrix}. \tag{19}
$$

Therefore, the matrix on the left hand side of (19) is positive definite and the system of equations has a unique solution, i. e. the COL predictor exists and is unique. Notice that the system (19) coincides with the system of linear equations for the weights in simple kriging, see e. g. [19, p. 24].

In the sub-Gaussian case, the system of linear equations (22) for the weights of the COL predictor coincides with (19) due to (13).

Remark 2. The fact that this method is well tractable makes it attractive for applications. Nevertheless it cannot cover all α-stable random fields because sub-Gaussian random fields have the property of long dependence (see [18, Proposition 4.7.4]), so they are a very particular case of stable fields.

Theorem 10. *If the covariance function C is continuous, then the COL predictor for Gaussian and sub-Gaussian random fields is continuous.*

If $t_0 \neq t_i$ for $i = 1, \ldots, n$, the Maximum-Likelihood (ML) predictor $\widehat{X(t_0)}$ of $X(t_0)$ is the solution of the optimization problem

$$
f_{X(t_0)|X(t_1)=x_1,\ldots,X(t_n)=x_n}(x) \; \to \; \max_{x \in \mathbb{R}}.
$$

If $t_0 = t_i$ for some $i = 1, \ldots, n$, we have

$$
\mathbb{P}(X(t_0) = x_i | X(t_1) = x_1, \ldots, X(t_i) = x_i, \ldots, X(t_n) = x_n) = 1 \tag{20}
$$

such that the ML predictor is $\widehat{X(t_0)} = X(t_i)$. Closed formulas for the ML predictor for sub-Gaussian random fields are provided by [16].

Theorem 11. *For Gaussian and sub-Gaussian random fields, the ML predictor is equal to the COL and LSL predictor.*

There exists an example for which the COL predictor is distinct from the LSL predictor.

Example 1. Let X be the $S\alpha S$ Lévy motion defined by $X(t) = \int_0^\infty \mathbb{1}(x \leq t)M(dx)$, where M is an $S\alpha S$ random measure with Lebesgue control measure. Let $t_0 = 3/4$ and $t_1 = 1$. Then the optimization problem for the LSL predictor is

$$
\begin{aligned}
\sigma^\alpha_{\widehat{X(t_0)}-X(t_0)} &= \int_0^{3/4} |1 - \lambda_1|^\alpha dx + \int_{3/4}^1 |\lambda_1|^\alpha dx \\
&= \frac{3}{4}|1 - \lambda_1|^\alpha + \frac{1}{4}|\lambda_1|^\alpha \; \to \; \min_{\lambda_1}.
\end{aligned}
$$

The last equation implies that the minimum is attained for some $\lambda_1 \in [0, 1]$ and we can therefore replace the absolute values by brackets. By setting the derivative of the last equation with respect to λ_1 equal to zero, we obtain the LSL predictor

$$
\widehat{X(t_0)} = \frac{1}{1 + (1/3)^{1/(\alpha-1)}} X(t_1).
$$

The COL predictor is

$$
\widehat{X(t_0)} = \frac{[X(t_0), X(t_1)]_\alpha}{[X(t_1), X(t_1)]_\alpha} X(t_1) = \frac{3}{4} X(t_1)
$$

and thus differs from the LSL predictor except for $\alpha = 2$.

3.2.2. Properties of the COL predictor for moving averages

In this subsection, we restrict our setting to moving average random fields of the form

$$X(t) = \int_{\mathbb{R}^d} f(t - x) M(dx), \quad t \in \mathbb{R}^d, \tag{21}$$

where M is an α-stable random measure with Lebesgue control measure. This restriction allows for imposing conditions on the kernel functions that guarantee the existence, uniqueness and continuity of the COL predictor.

We now state some properties of such moving average random fields. Let $h \in \mathbb{R}^d$. Since X is stationary, it holds $[X(h), X(0)]_\alpha = [X(t + h), X(t)]_\alpha$ for each $t \in \mathbb{R}^d$. Therefore, the covariation function of X is a function of just one variable and we use the notation $\kappa(h) = [X(h), X(0)]_\alpha$.

Lemma 12. *If the kernel function $f : \mathbb{R}^d \to \mathbb{R}_+$ of the moving average (21) is positive semi-definite, then the covariation function κ is positive semi-definite. If $f : \mathbb{R}^d \to \mathbb{R}_+$ is positive definite and positive on a set with positive Lebesgue measure, then κ is positive definite.*

It is clear that the COL predictor is unique if the covariation function is positive definite since the matrix of the system of linear equations

$$\begin{pmatrix} \kappa(0) & \cdots & \kappa(t_n - t_1) \\ \vdots & \ddots & \vdots \\ \kappa(t_n - t_1) & \cdots & \kappa(0) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} \kappa(t_0 - t_1) \\ \vdots \\ \kappa(t_0 - t_n) \end{pmatrix} \tag{22}$$

is positive definite. An example of an α-stable moving average with a positive definite kernel function is the $S\alpha S$ Ornstein-Uhlenbeck process with

$$X(t) = \int_{\mathbb{R}} e^{-\lambda(t-x)} \mathbb{1}(t - x \geq 0) M(dx), \quad t \in \mathbb{R},$$

for some $\lambda > 0$. If $t_1 < t_2 < \ldots < t_n < t_0$, then the regression of $X(t_0)$ on $(X(t_1), \ldots, X(t_n))^\mathsf{T}$ is linear (cf. [18, p. 138]), i. e. $\widehat{X(t_0)} = e^{-\lambda(t_0 - t_n)} X(t_n)$.

Theorem 13. *If the covariation function is positive definite and continuous, then the COL predictor is continuous.*

The proof is analogous to the one of Theorem 10. For example, all continuous kernel functions with compact support yield a continuous covariation function by the dominated convergence theorem.

3.3. Maximization of Covariation Linear (MCL) predictor

The idea of this method is to maximize the covariation

$$\left[\widehat{X(t_0)}, X(t_0) \right]_\alpha = \sum_{i=1}^{n} \lambda_i \left[X(t_i), X(t_0) \right]_\alpha$$

between the linear predictor $\widehat{X(t_0)}$ and the theoretical value $X(t_0)$. In general, if we do not impose any further restrictions and if $[X(t_i), X(t_0)]_\alpha \neq 0$ for some $i \in \{1, \ldots, n\}$, then the maximization problem

$$\sum_{i=1}^{n} \lambda_i \left[X(t_i), X(t_0) \right]_\alpha \to \max_{\lambda_1, \ldots, \lambda_n} \tag{23}$$

obviously does not have a solution. A reasonable side condition is

$$\sigma_{\widehat{X(t_0)}} = \sigma_{X(t_0)}$$

which guarantees $\widehat{X(t_0)} \overset{d}{=} X(t_0)$ for $S\alpha S$ random fields X. The optimization problem becomes

$$\begin{cases} \left[\widehat{X(t_0)}, X(t_0) \right]_\alpha \to \max\limits_{\lambda_1, \ldots, \lambda_n}, \\ \sigma_{\widehat{X(t_0)}} = \sigma_{X(t_0)}. \end{cases} \tag{24}$$

The predictor $\widehat{X(t_0)} = \sum_{i=1}^{n} \lambda_i X(t_i)$ whose weights $\lambda_1, \ldots, \lambda_n$ are a solution of the maximization problem (24) is called *Maximization of Covariation Linear (MCL) predictor*.

3.3.1. Properties of the MCL predictor

The following theorem provides sufficient conditions for the existence, uniqueness, exactness and continuity of the MCL predictor.

Theorem 14. *Assume that the random vector* $(X(t_1), \ldots, X(t_n))^\mathsf{T}$ *is full-dimensional and* $t_0 \in \mathbb{R}^d$. *Then the following holds:*

1. *The MCL predictor exists.*
2. *If additionally* $\kappa(t_i, t_0) \neq 0$ *for some* $i \in \{1, \ldots, n\}$, *then the MCL predictor at* t_0 *is unique.*
3. *If the MCL predictor is unique, then it is exact.*
4. *If* $\kappa(t_i, t_0) \neq 0$ *for some* $i \in \{1, \ldots, n\}$ *and* κ *is continuous, then the MCL predictor is continuous in* t_0.

Remark 3. We have already seen in Section 3.2.2 that for moving averages, a sufficient condition for the continuity of the covariation function is that the kernel function is continuous and has a compact support.

4. Numerical results

In this section, we apply the extrapolation methods to a sub-Gaussian random field and an $S\alpha S$ Lévy motion on the square $[0, 1]^2$. For this, we generate 1000 realizations of both random fields on an equidistant grid of 100×100 points and extrapolate each realization by using the values $x(t_1), \ldots, x(t_9)$ at the locations $t_1 = (0.2, 0.2)^\mathsf{T}$, $t_2 = (0.2, 0.5)^\mathsf{T}$, $t_3 = (0.2, 0.8)^\mathsf{T}$, $t_4 = (0.5, 0.2)^\mathsf{T}$, $t_5 = (0.5, 0.5)^\mathsf{T}$, $t_6 = (0.5, 0.8)^\mathsf{T}$, $t_7 = (0.8, 0.2)^\mathsf{T}$, $t_8 = (0.8, 0.5)^\mathsf{T}$ and $t_9 = (0.8, 0.8)^\mathsf{T}$.

The sub-Gaussian random field $X = \{A^{1/2}G(t), t \in [0, 1]^2\}$ has the stability index $\alpha = 1.5$, where G is a stationary Gaussian random field with Gaussian covariance function $C(h) = 7 \exp\{-(h/0.1)^2\}$, $h \geq 0$. Recall that in this case, the LSL, COL and ML predictors coincide by Theorem 11. Due to the specific structure of sub-Gaussian random fields, we also provide a conditional simulation (CS) of X, i. e. a simulation of X that honours the conditions $X(t_i) = x(t_i)$ for $i = 1, \ldots, 9$, see [12]. The conditional simulation can be seen as another extrapolation method and carried out straightforward by conditionally simulating the Gaussian part G of X and scaling it appropriately with a realization of $A^{1/2}$. A conditional simulation algorithm for Gaussian random fields is given in [12]. For a simulation algorithm for A, see [5]. Figure 1 shows a realization of X and the extrapolations based on the LSL (COL, ML), MCL and CS methods.

Table 1: Summary statistics for the deviations $X(g) - \widehat{X(g)}$ (sub-Gaussian random field)

Method	5%-Quantile	1st Quartile	Median	Mean	3rd Quartile	95%-Quantile
LSL (COL, ML)	-1.5451	-0.4446	0.0018	0.0070	0.4503	1.5363
MCL	-1.8204	-0.4899	0.0046	-0.0188	0.5016	1.7580
CS	-2.7523	-0.5837	0.0058	0.0057	0.5985	2.7262

In order to assess the performance of the predictors, we calculated the deviations $X(g) - \widehat{X(g)}$ of the extrapolated values from the simulated ones at each grid point g of the 1000 realizations of the sub-Gaussian random field. The summary statistics of the deviations are shown in Table 1. As one can see, there are only marginal differences between the quantiles and the mean of the deviations for the three methods, the LSL (COL, ML) and MCL method performing slightly better than the CS method.

Consider now the two-dimensional $S\alpha S$ Lévy motion Y with stability index $\alpha = 1.5$ on the square $[0, 1]^2$ defined by $Y(t) = \int_{[0,1]^2} \mathbb{1}\{x_1 \leq t_1, x_2 \leq t_2\} M(dx)$ for $t \in [0, 1]^2$, where M is an $S\alpha S$ random measure with Lebesgue control measure. Figure 2 shows a realization of X and the extrapolations based on the LSL, COL and MCL methods.

Again, we investigate the performance of the extrapolation methods by calculating the summary statistics of the deviations of the extrapolated values from the simulated ones. Table 2 shows the corresponding results. As for the sub-Gaussian random field X, no method clearly outperforms the others.
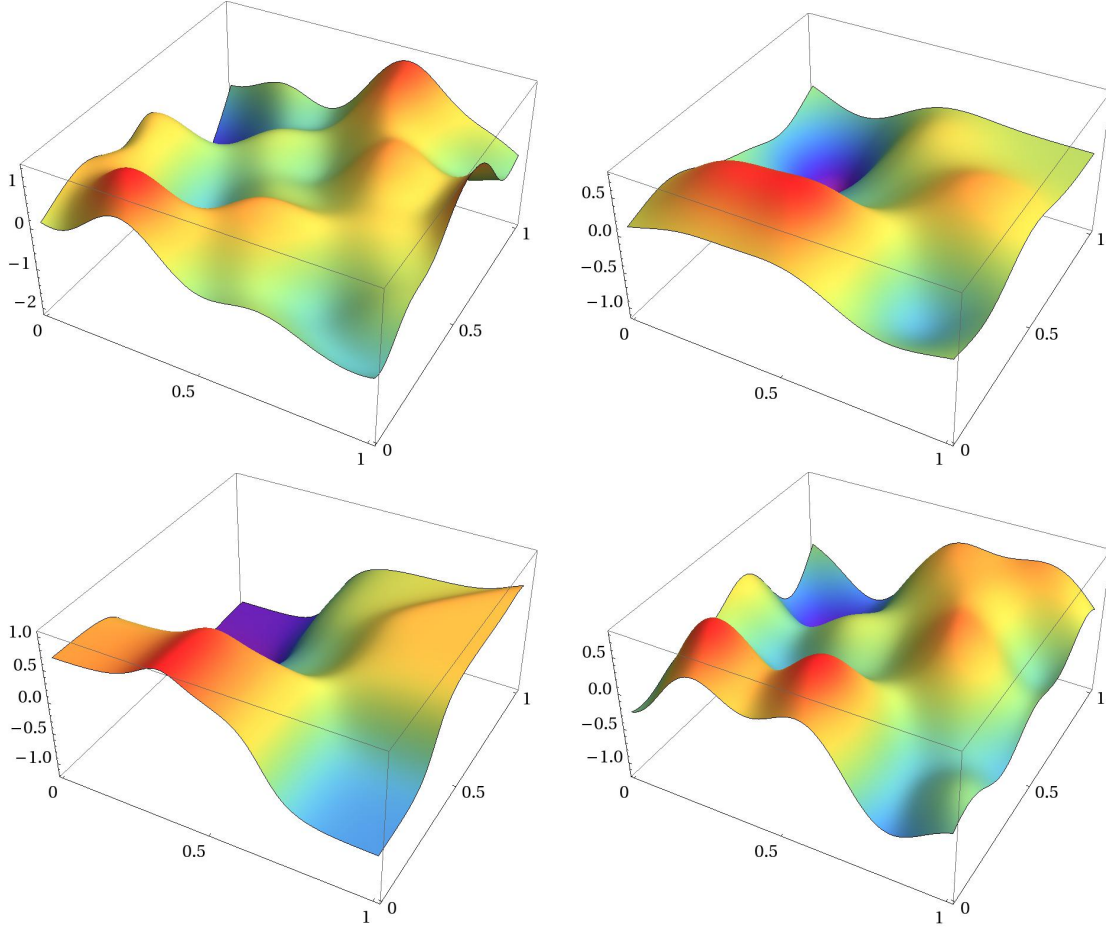
Figure 1: Realization of the sub-Gaussian random field (top left) and the extrapolations based on the LSL (COL, ML) method (top right), the MCL method (bottom left) and the CS method (bottom right).

Table 2: Summary statistics for the deviations $X(g) - \widehat{X(g)}$ ($S\alpha S$ Lévy motion)

Method	5%-Quantile	1st Quartile	Median	Mean	3rd Quartile	95%-Quantile
LSL	-0.5170	-0.1246	0.0000	-0.0071	0.1226	0.5045
COL	-0.5263	-0.1289	0.0002	-0.0061	0.1266	0.5137
MCL	-0.6093	-0.1455	-0.0007	0.0283	0.1407	0.5895

5. Discussion and open problems

For sub-Gaussian random fields, the LSL and COL methods can be efficiently implemented and easily applied to larger data sets of at least 1000 sample points since efficient algorithms to solve linear systems of equations are available. For other kinds of α-stable random fields, the COL predictor can still be implemented efficiently for large data sets since only a system of linear equations has to be solved. In the LSL and MCL methods, a function whose dimension equals the number of sample points needs to be numerically minimized and maximized, respectively, impeding the application of these two methods to large data sets.

We notice that we have proven the properties of the LSL, COL and MCL predictors for α-stable random fields
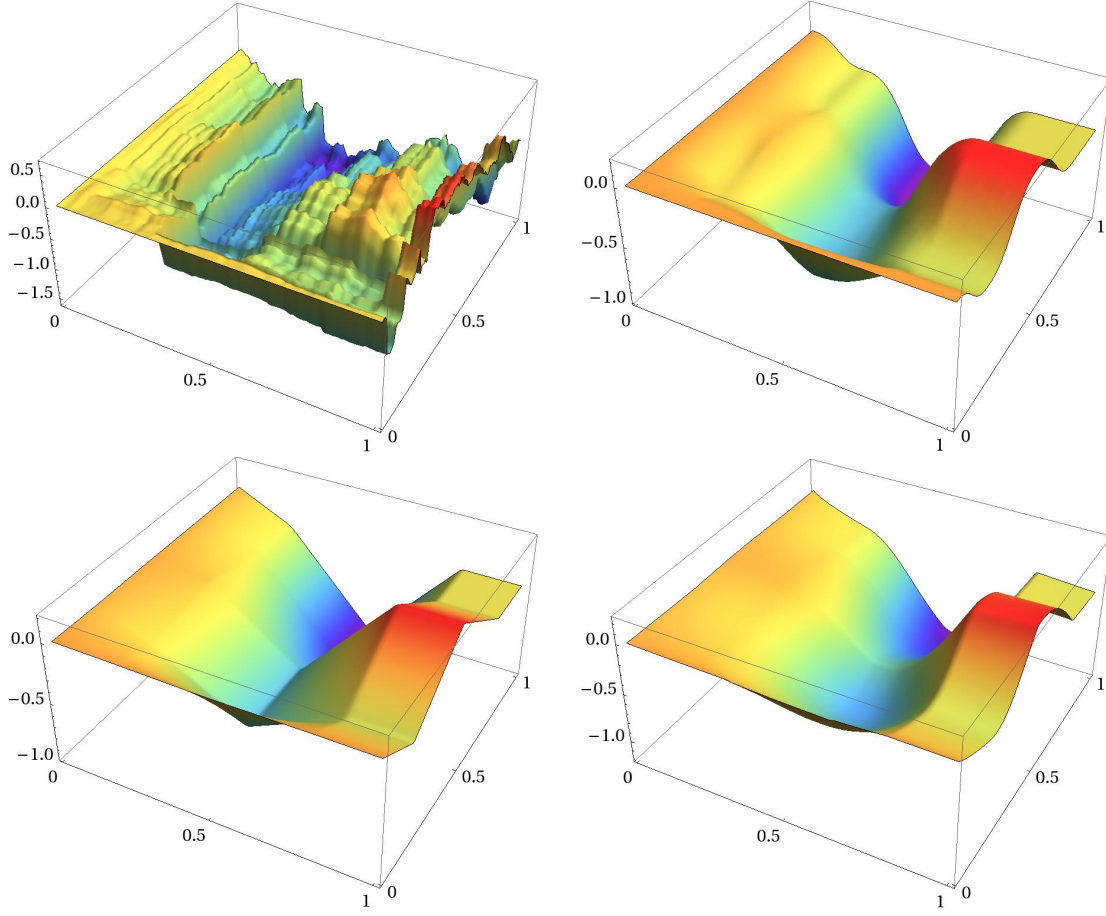
Figure 2: Realization of the Lévy stable motion (top left) and the extrapolations based on the LSL method (top right), the COL method (bottom left) and the MCL method (bottom right).

with integral representation (6) and $\alpha > 1$. All of these methods do not involve the skewness of the random field X. If X is $S\alpha S$, then obviously so is $\hat{X}$ as it is a linear combination of $S\alpha S$ random variables. If the skewness function β of X is not zero, then easy calculations show that the skewness parameter of $\hat{X}(t) = \sum_{i=1}^{n} \lambda_i X(t_i)$ is given by

$$\frac{\int_E \left(\sum_{i=1}^{n} \lambda_i f_{t_i}(x) \right)^{\langle \alpha \rangle} \beta(x) m(dx)}{\int_E \left| \sum_{i=1}^{n} \lambda_i f_{t_i}(x) \right|^{\alpha} m(dx)},$$

cf. [18, Property 3.2.2].

The ML and CS predictors can be immediately extended to the case $\alpha \leq 1$, but they are only available for Gaussian and sub-Gaussian random fields. Also, the minimization problem (15) in the LSL method can be considered for $\alpha \leq 1$, but except for the case $\alpha = 1$, the objective function will change from convexity to concavity in $\lambda = (\lambda_1, \ldots, \lambda_n)^\top$ and it is not clear whether the properties of the LSL predictor still hold. The development of extrapolation methods for $\alpha \leq 1$ is thus an interesting problem for future research.

Another interesting open problem is a characterization of the covariation function which has been touched upon only slightly in Lemma 12.

Acknowledgement

Appendix A. Proofs

Proof of Lemma 1. Let $supp(\Gamma)$ denote the support of Γ and assume that $supp(\Gamma) \subset \mathbb{S}^{d-1} \cap G_{c_1,\dots,c_d}$ for some $(d-1)$-dimensional subspace $G_{c_1,\dots,c_d} := \{(x_1,\dots,x_d)^\top \in \mathbb{R}^d : c_1 x_1 + \dots + c_d x_d = 0\}$ with $c_1,\dots,c_d \in \mathbb{R}$. Let us consider $\sum_{i=1}^d c_i X_i$. Using [18, Example 2.3.4], we obtain

$$\sigma^\alpha_{\sum_{i=1}^d c_i X_i} = \int_{\mathbb{S}^{d-1}} \left| \sum_{i=1}^d c_i s_i \right|^\alpha \Gamma(ds) = \int_{\mathbb{S}^{d-1} \cap G_{c_1,\dots,c_d}} \left| \sum_{i=1}^d c_i s_i \right|^\alpha \Gamma(ds) = 0$$

Since $\mu = 0$, we have $\sum_{i=1}^d c_i X_i = 0$ almost surely.

Vice versa, if $\sum_{i=1}^d c_i X_i = 0$ almost surely, then $\sigma^\alpha_{\sum_{i=1}^d c_i X_i} = \int_{\mathbb{S}^{d-1}} \left| \sum_{i=1}^d c_i s_i \right|^\alpha \Gamma(ds) = 0$. Thus Γ is concentrated on the great sub-sphere $G_{c_1,\dots,c_d} := \{(s_1,\dots,s_d)^\top \in \mathbb{S}^{d-1} : \sum_{i=1}^d c_i s_i = 0\}$ of $\mathbb{S}^{d-1}$. $\qquad\square$

Proof of Lemma 3. Let $E_+ := \{x \in E : f_1(x)^2 + f_2(x)^2 > 0\}$ and $g_j : E \to \mathbb{R}$ be defined by

$$g_j(x) := \frac{f_j(x)}{(f_1(x)^2 + f_2(x)^2)^{1/2}}, \quad x \in E, \quad j = 1, 2.$$

Then we have for any Borel set $A \subset \mathbb{S}^1$

$$\Gamma(A) = \int_{g^{-1}(A)} \frac{1 + \beta(x)}{2} \left(f_1(x)^2 + f_2(x)^2 \right)^{\alpha/2} m(dx) + \int_{g^{-1}(-A)} \frac{1 - \beta(x)}{2} \left(f_1(x)^2 + f_2(x)^2 \right)^{\alpha/2} m(dx),$$

where $g^{-1}(A) := \{x \in E_+ : (g_1(x), g_2(x))^\top \in A\}$, see [18, pp. 115]. Let $S^- = \{(s_1, s_2)^\top \in \mathbb{S}^1 : s_1 s_2 < 0\}$. It holds $S^- = -S^-$. The random vector $(X, Y)^\top$ is associated if and only if

$$\Gamma(S^-) = \int_{g^{-1}(S^-)} \left(f_1(x)^2 + f_2(x)^2 \right)^{\alpha/2} m(dx) = 0,$$

see [18, Theorem 4.6.1]. This is equivalent to $m(g^{-1}(S^-)) = 0$, that is $f_1 f_2 \geq 0$ m-almost everywhere. The proof for negative association is analogous by using [18, Theorem 4.6.3]. $\qquad\square$

Proof of Corollary 4. By [18, Theorem 3.5.6], $(X, Y)^\top$ has an integral representation

$$(X, Y)^\top \overset{d}{=} \left(\int_E f_1(x) M(dx), \int_E f_2(x) M(dx) \right)^\top,$$

where M is an α-stable random measure with control measure m and skewness intensity function β. We define $E^* := \{x \in E : f_1(x) f_2(x) \geq 0\}$, $X_1 := \int_E f_1(x) \mathbb{1}_{E^*}(x) M(dx)$, $Y_1 := \int_E f_2(x) \mathbb{1}_{E^*}(x) M(dx)$, $X_2 := X - X_1$ and $Y_2 := Y - Y_1$. The dependence properties of X_1, X_2, Y_1 and Y_2 follow immediately from Lemma 3 and the fact that an α-stable random vector $(U, V)^\top = \left(\int_E g_1(x) M(dx), \int_E g_2(x) M(dx) \right)^\top$ has independent components if and only if $g_1 g_2 = 0$ m-almost everywhere, see [18, Theorem 3.5.3].. $\qquad\square$

Proof of Lemma 6. By [18, Property 1.2.17], we have $\mathbb{E}|\lambda X + Y|^p = \mathbb{E}|\xi|^p \sigma^p_{\lambda X + Y}$ for $\lambda \in \mathbb{R}$, where $\xi \sim S_\alpha(1, \beta_{\lambda X + Y}, 0)$ and $\beta_{\lambda X + Y} = \int_{\mathbb{S}^1} (\lambda s_1 + s_2)^{<\alpha>} \Gamma(ds) \big/ \int_{\mathbb{S}^1} |\lambda s_1 + s_2|^\alpha \Gamma(ds)$. Taking derivatives on both sides, we get

$$\begin{aligned}
\frac{d}{d\lambda} \mathbb{E}|\lambda X + Y|^p \bigg|_{\lambda=0} &= p\mathbb{E}\left(XY^{<p-1>} \right), \\
\frac{d}{d\lambda} \mathbb{E}|\xi|^p \sigma^p_{\lambda X + Y} \bigg|_{\lambda=0} &= p\sigma_Y^{p-\alpha} [X, Y]_\alpha \, \mathbb{E}|\xi|^p \big|_{\lambda=0} + \sigma_Y^p \frac{d}{d\lambda} \mathbb{E}|\xi|^p \bigg|_{\lambda=0}.
\end{aligned} \tag{A.1}$$

A closed formula for $\mathbb{E}|\xi|^p$ is given in [18, p. 18]. After some cumbersome calculations, we obtain

$$
\begin{aligned}
\mathbb{E}|\xi|^p\big|_{\lambda=0} &= \mathbb{E}|Y|^p/\sigma_Y^p, \\
\frac{d}{d\lambda}\mathbb{E}|\xi|^p\bigg|_{\lambda=0} &= p\frac{\mathbb{E}|Y|^p}{\sigma_Y^{p+\alpha}}\cdot c\cdot\left((X,Y)_\alpha - \beta_Y[X,Y]_\alpha\right).
\end{aligned}
$$

Plugging these formulas in (A.1) and equating both derivatives yields the assertion. $\qquad\square$

Appendix A.1. Proofs for the LSL predictor

Proof of Lemma 7. Let $\lambda_0 := -1$. Then we get for $j = 1,\dots,n$

$$
\sigma_{\widehat{X(t_0)}-X(t_0)}^\alpha = \sigma_{\sum_{i=0}^n \lambda_i X(t_j)}^\alpha = \int_{\mathbb{S}^n}\left|\sum_{i=0}^n \lambda_i s_i\right|^\alpha \Gamma(ds),
$$

where Γ is the spectral measure of the random vector $(X(t_0),\dots,X(t_n))^\top$, cf. [18, formula (2.3.3)]. It is easy to see that for $j = 1,\dots,n$, the partial derivative $\frac{d}{d\lambda_j}\left|\sum_{i=0}^n \lambda_i s_i\right|^\alpha$ (as a function of λ_j) is uniformly bounded in $(s_0,\dots,s_n)^\top \in \mathbb{S}^n$ if $\alpha > 1$. Then since Γ is a finite measure, it follows from the dominated convergence theorem that

$$
\begin{aligned}
\frac{\partial}{\partial\lambda_j}\sigma_{\widehat{X(t_0)}-X(t_0)}^\alpha &= \int_{\mathbb{S}^n}\frac{\partial}{\partial\lambda_j}\left|\sum_{i=0}^n \lambda_i s_i\right|^\alpha \Gamma(ds) \\
&= \alpha\int_{\mathbb{S}^n} s_j\left(\sum_{i=0}^n \lambda_i s_i\right)^{<\alpha-1>}\Gamma(ds) = \alpha\left[X(t_j),\sum_{i=1}^n \lambda_i X(t_i) - X(t_0)\right]_\alpha,
\end{aligned}
$$

see [18, Lemma 2.7.5]. $\qquad\square$

Proof of Theorem 8. Assertion 1 follows from [6, Theorem 1.1, p. 59], the discussion thereafter and Lemma 2. The exactness of the predictor is obvious. We now prove the second assertion. Since X is stochastically continuous, we have by (7) and (8)

$$
\|f_s - f_t\|_{L^\alpha(E,m)} \to 0, \quad s \to t,
$$

where $\|\cdot\|_{L^\alpha(E,m)}$ denotes the usual norm of $L^\alpha(E,m)$. Let $\hat{f}_t = \sum_{i=1}^n \lambda_i(t)f_{t_i}$ for each $t \in \mathbb{R}^d$. Then it holds

$$
\left\|\hat{f}_s - \hat{f}_t\right\|_{L^\alpha(E,m)} = \left\|\sum_{i=1}^n (\lambda_i(s) - \lambda_i(t))f_{t_i}\right\|_{L^\alpha(E,m)} \to 0, \quad s \to t, \tag{A.2}
$$

see [6, Theorem 1.2, p. 60].

Assume that $\lambda_i(s) \nrightarrow \lambda_i(t)$ as $s \to t$ for some $i \in \{1,\dots,n\}$. Then $\sqrt{(\lambda_1(s) - \lambda_1(t))^2 + \cdots + (\lambda_n(s) - \lambda_n(t))^2} \nrightarrow 0$ as $s \to t$. Since

$$
\left|\frac{\lambda_i(s) - \lambda_i(t)}{\sqrt{(\lambda_1(s) - \lambda_1(t))^2 + \cdots + (\lambda_n(s) - \lambda_n(t))^2}}\right| \le 1, \quad i = 1,\dots,n,
$$

there exists a sequence $\{s_k\}_{k\in\mathbb{N}}$ with $s_k \to t$ as $k \to \infty$ such that

$$
\gamma_i^{(k)} := \frac{\lambda_i(s_k) - \lambda_i(t)}{\sqrt{(\lambda_1(s_k) - \lambda_1(t))^2 + \cdots + (\lambda_n(s_k) - \lambda_n(t))^2}} \to \gamma_i, \quad k \to \infty, \quad i = 1,\dots,n.
$$

Notice that $(\gamma_1^{(k)})^2 = 1 - \sum_{i=2}^n (\gamma_i^{(k)})^2$ for each $k \in \mathbb{N}$ such that $\gamma_i \neq 0$ for some $i \in \{1,\dots,n\}$. By (A.2), we have $\|\sum_{i=1}^n \gamma_i f_{t_i}\|_{L^\alpha(E,m)} = 0$ which implies $\sum_{i=1}^n \gamma_i f_{t_i} = 0$ m-almost everywhere. By Lemma 2, this is a contradiction to the the full-dimensionality of $(X(t_1),\dots,X(t_n))^\top$. $\qquad\square$

Appendix A.2. Proofs for the COL predictor

Proof of Theorem 10. Let $t_0 \in \mathbb{R}^d$. The matrix on the left hand side of equation (19) is a bijective continuous linear operator. By the open mapping theorem, the inverse operator is continuous. Since the covariance function is continuous, the weights $\lambda_1, \ldots, \lambda_n$ (considered as real-valued functions on $\mathbb{R}^d$) are continuous in t_0. $\square$

Proof of Theorem 11. Let A be a lower triangular $(n + 1) \times (n + 1)$-matrix with the property

$$A\Omega_{t_1,\ldots,t_n,t_0}A^\mathsf{T} = 2I, \tag{A.3}$$

where I denotes the identity matrix and $\Omega_{t_1,\ldots,t_n,t_0}$ is the covariance matrix of the Gaussian part of the random vector $(X(t_1), \ldots, X(t_n), X(t_0))^\mathsf{T}$. Then the ML predictor for $X(t_0)$ is given by

$$\widehat{X(t_0)} = \begin{cases} \sum_{i=1}^n \frac{-a_{n+1,i}}{a_{n+1,n+1}} X(t_i), & t_0 \neq t_i \text{ for all } i \in \{1, \ldots, n\}, \\ X(t_i), & t_0 = t_i \text{ for some } i \in \{1, \ldots, n\}. \end{cases} \tag{A.4}$$

where $a_{i,j}$ denotes the entry in the i-th row and j-th column of A, see [16]. Let C be the covariance function of the Gaussian part G of the sub-Gaussian random field X. We have $\Omega_{t_1,\ldots,t_n,t_0} = 2(A^\mathsf{T}A)^{-1}$ and $A = B^{-1}$, where B is the lower triangular matrix of the Cholesky decomposition of $\Omega_{t_1,\ldots,t_n,t_0}$, i. e. $\Omega_{t_1,\ldots,t_n,t_0} = 2BB^\mathsf{T}$. Notice that the last column of the matrix $-1/a_{n+1,n+1} \cdot A^\mathsf{T}$ consists of the coefficients $-a_{n+1,1}/a_{n+1,n+1}, \ldots, -a_{n+1,n}/a_{n+1,n+1}$ of the ML predictor and the number -1. Let $\lambda_i = -a_{n+1,i}/a_{n+1,n+1}$ for $i = 1, \ldots, n$. We have

$$\frac{1}{2a_{n+1,n+1}}\Omega_{t_1,\ldots,t_n,t_0}A^\mathsf{T} = \frac{1}{a_{n+1,n+1}}(A^\mathsf{T}A)^{-1}A^\mathsf{T} = \frac{1}{a_{n+1,n+1}}A^{-1}(A^\mathsf{T})^{-1}A^\mathsf{T}$$

$$= \frac{1}{a_{n+1,n+1}}A^{-1} = \frac{1}{a_{n+1,n+1}}B = \begin{pmatrix} x_{1,1} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,n} & 0 \\ x_{n+1,1} & \cdots & x_{n+1,n} & x_{n+1,n+1} \end{pmatrix}$$

for some real numbers $x_{1,1}, \ldots, x_{n+1,n+1}$. By only considering the last column of the left and right hand side of the last equation yields

$$\begin{pmatrix} C(0) & \cdots & C(t_n - t_1) \\ \vdots & \ddots & \vdots \\ C(t_n - t_1) & \cdots & C(0) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} - \begin{pmatrix} C(t_0 - t_1) \\ \vdots \\ C(t_0 - t_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

which is the system of linear equations that has to be solved for the COL predictor. Since the COL predictor is unique, it must be equal to the ML predictor.

We now show that the COL predictor is equal to the LSL predictor. Let $X = \{A^{1/2}G(t), t \in \mathbb{R}^d\}$ be a sub-Gaussian random field. We have $\widehat{X(t_0)} - X(t_0) = A^{1/2} \sum_{i=0}^n \lambda_i G(t_i)$, with $\lambda_0 = -1$ and

$$\mathrm{Var}\left(\sum_{i=0}^n \lambda_i G(t_i) \right) = \sum_{i,j=0}^n \lambda_i \lambda_j C(t_i - t_j).$$

With [18, Proposition 1.3.1], we get

$$\sigma_{\widehat{X(t_0)}-X(t_0)}^\alpha = \left(\frac{1}{2} \sum_{i,j=0}^n \lambda_i \lambda_j C(t_i - t_j) \right)^{\alpha/2} \to \min_{\lambda_1,\ldots,\lambda_n}$$

which is equivalent to

$$\sum_{i,j=0}^n \lambda_i \lambda_j C(t_i - t_j) \to \min_{\lambda_1,\ldots,\lambda_n}.$$

Taking partial derivatives and setting them equal to zero shows that the LSL predictor is equal to the COL predictor. The calculations for Gaussian random fields are analogous. $\square$

Proof of Lemma 12. Let $n \in \mathbb{N}$, $z_1, \ldots, z_n \in \mathbb{R}$ and $t_1, \ldots, t_n \in \mathbb{R}^d$. It holds

$$\sum_{i,j=1}^n \kappa(t_i - t_j)z_i z_j = \left[\sum_{i,j=1}^n X(t_i - t_j)z_i z_j, X(0)\right]_\alpha$$

$$= \int_{\mathbb{R}^d} \sum_{i,j=1}^n f(t_i - t_j - x)z_i z_j f^{\langle \alpha-1\rangle}(-x)dx \geq 0$$

if f is positive semi-definite and $f(x) \geq 0$ for all $x \in \mathbb{R}^d$. We consider $(z_1, \ldots, z_n)^\mathsf{T} \neq (0, \ldots, 0)^\mathsf{T}$ for the positive definiteness of κ. Let A be the set where f is positive with positive Lebesgue measure. Then $-x \in -A$ if and only if $x \in A$ and

$$\sum_{i,j=1}^n \kappa(t_i - t_j)z_i z_j \geq \int_{-A} \sum_{i,j=1}^n f(t_i - t_j - x)z_i z_j f^{\langle\alpha-1\rangle}(-x)dx > 0.$$

$\square$

Appendix A.3. Proofs for the MCL predictor

Proof of Theorem 14. 1. Let $C := \{(\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n : \sigma_{\sum_{i=1}^n \lambda_i X(t_i)} = \sigma_{X(t_0)}\}$. Let $\lambda_2 = \ldots = \lambda_n = 0$. Then $\sigma_{\sum_{i=1}^n \lambda_i X(t_i)} = \sigma_{\lambda_1 X(t_1)} = |\lambda_1|\sigma_{X(t_1)}$ such that $\lambda_1 = \pm\sigma_{X(t_0)}/\sigma_{X(t_1)}$. Therefore, the set C is non-empty. Since $L^\alpha(E, m)$ is a linear space, $\sigma_{\sum_{i=1}^n \lambda_i X(t_i)}$ is a continuous function in $\lambda = (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n$ by Lebesgue's theorem on dominated convergence. Thus the set C is closed. By Lemma 2, the functions f_{t_i}, $i = 1, \ldots, n$, are linearly independent. Due to this, $\sigma_{\sum_{i=1}^n \lambda_i X(t_i)} = \int_E |\sum_{i=1}^n \lambda_i f_{t_i}(x)|^\alpha m(dx)$ tends to infinity as $\|\lambda\|_2 \to \infty$ so that C is bounded and hence compact. Since the objective function is continuous, there exists a solution of the optimization problem.

2. Recall the optimization problem

$$\begin{cases} \left[\widehat{X(t_0)}, X(t_0)\right]_\alpha \to \max_{\lambda_1, \ldots, \lambda_n}, \\ \sigma_{\widehat{X(t_0)}} = \sigma_{X(t_0)}. \end{cases} \tag{A.5}$$

Define the function $\Psi : \mathbb{R}^n \to \mathbb{R}$ by

$$\Psi(\lambda) := \sigma_{\widehat{X(t_0)}} = \left\|\sum_{i=1}^n \lambda_i f_{t_i}\right\|_\alpha, \quad \lambda = (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n.$$

It is obvious that Ψ is continuous. By Minkowski's inequality, Ψ is convex and strictly convex on the set of admissible points given by $\Lambda := \{\lambda = (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n : \Psi(\lambda) = \sigma_{X(t_0)}\}$, that is

$$\Psi(\beta\lambda^{(1)} + (1 - \beta)\lambda^{(2)}) < \beta\Psi(\lambda^{(1)}) + (1 - \beta)\Psi(\lambda^{(2)})$$

for all $\beta \in (0, 1)$ and $\lambda^{(1)}, \lambda^{(2)} \in \Lambda$ with $\lambda^{(1)} \neq \lambda^{(2)}$.
Indeed, let $\lambda^{(1)}, \lambda^{(2)} \in \Lambda$ with $\lambda^{(1)} \neq \lambda^{(2)}$ and $\beta \in (0, 1)$. We have

$$\Psi(\beta\lambda^{(1)} + (1 - \beta)\lambda^{(2)}) = \left\|\beta\sum_{i=1}^n \lambda_i^{(1)} f_{t_i} + (1 - \beta)\sum_{i=1}^n \lambda_i^{(2)} f_{t_i}\right\|_\alpha.$$

Assume that

$$\beta\sum_{i=1}^n \lambda_i^{(1)} f_{t_i}(x) = \gamma(1 - \beta)\sum_{i=1}^n \lambda_i^{(2)} f_{t_i}(x) \quad m - a.e. \tag{A.6}$$

for some $\gamma \geq 0$. By Lemma 2, this contradicts the full-dimensionality of the vector $(X(t_1), \ldots, X(t_n))^\mathsf{T}$. Analogously, by exchanging β and $1 - \beta$, we have

$$(1 - \beta)\sum_{i=1}^n \lambda_i^{(2)} f_{t_i}(x) \neq \gamma\beta\sum_{i=1}^n \lambda_i^{(1)} f_{t_i}(x) \quad m - a.e.$$

16

for all $\gamma \geq 0$. The Minkowski inequality yields $\|g + h\|_\alpha \leq \|g\|_\alpha + \|h\|_\alpha$ for all $g, h \in L^\alpha(E, m)$ and equality holds if and only if $g = \gamma h$ or $h = \gamma g$ m-almost everywhere for some $\gamma \geq 0$. We conclude that

$$
\begin{aligned}
\Psi(\beta\lambda^{(1)} + (1-\beta)\lambda^{(2)}) &= \left\| \beta \sum_{i=1}^n \lambda_i^{(1)} f_{t_i} + (1-\beta) \sum_{i=1}^n \lambda_i^{(2)} f_{t_i} \right\|_\alpha \\
&< \beta \left\| \sum_{i=1}^n \lambda_i^{(1)} f_{t_i} \right\|_\alpha + (1-\beta) \left\| \sum_{i=1}^n \lambda_i^{(2)} f_{t_i} \right\|_\alpha = \beta\Psi(\lambda^{(1)}) + (1-\beta)\Psi(\lambda^{(2)})
\end{aligned}
$$

such that Ψ is strictly convex on Λ.

Consider now the set $B := \{\lambda \in \mathbb{R}^n : \Psi(\lambda) \leq \sigma_{X(t_0)}\}$. B is convex since Ψ is convex. Moreover, it is strictly convex which can be shown as follows. Let $\lambda \in \partial B$, where ∂B denotes the boundary of B. Then $\Psi(\lambda) = \sigma_{X(t_0)}$ because otherwise $\Psi(\lambda) < \sigma_{X(t_0)}$, and since Ψ is continuous, there exists an $\varepsilon > 0$ such that $\Psi(x) < \sigma_{X(t_0)}$ for all $x \in B_\lambda(\varepsilon) := \{x \in \mathbb{R}^n : \|x - \lambda\|_2 \leq \varepsilon\}$. Therefore, $B_\lambda(\varepsilon) \subset B$ which is a contradiction to $\lambda \in \partial B$. Let $\lambda^{(1)}, \lambda^{(2)} \in \partial B$ and $\beta \in (0, 1)$. We have already shown that

$$
\Psi(\beta\lambda^{(1)} + (1-\beta)\lambda^{(2)}) < \beta\Psi(\lambda^{(1)}) + (1-\beta)\Psi(\lambda^{(2)}) = \sigma_{X(t_0)},
$$

so $\beta\lambda^{(1)} + (1-\beta)\lambda^{(2)} \in \mathring{B}$, where $\mathring{B}$ denotes the interior of B. Therefore, B is strictly convex.

The strict convexity of B implies that for each $u \in \mathbb{R}^n$, the support set

$$
T(B, u) = \{y \in B : \langle y, u \rangle = h(B, u)\} \tag{A.7}
$$

consists of a singleton $\{x\}$, cf. the proof of [14, Theorem 4.5], where

$$
h(B, u) = \sup\{\langle x, u \rangle : x \in B\} \tag{A.8}
$$

is the support function of B. It is clear that $x \in \partial B$. Recall that the optimization problem (A.5) can be written as

$$
(P) \begin{cases} \langle \lambda, a \rangle \to \max \\ \lambda \in \Lambda \end{cases}
$$

where

$$
\begin{aligned}
a &:= ([X(t_1), X(t_0)]_\alpha, \ldots, [X(t_n), X(t_0)]_\alpha)^\mathsf{T}, \\
\Lambda &:= \{\lambda = (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n : \sigma_{\sum_{i=1}^n \lambda_i X(t_i)} = \sigma_{X(t_i)}\}.
\end{aligned}
$$

Consider the optimization problem

$$
(P^*) \begin{cases} \langle \lambda, a \rangle \to \max \\ \lambda \in B \end{cases}
$$

which has a unique solution $\lambda^* = T(B, a) \in \partial B$ by (A.7) and (A.8) since $T(B, a)$ consists of only one element. We have already seen that $\partial B \subset \Lambda$ such that λ^* is the unique solution of (P), too.

3. The Lagrange function $L : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ of the optimization problem (A.5) is given by

$$
L(\lambda, \gamma) = \sum_{i=1}^n \lambda_i [X(t_i), X(t_0)]_\alpha + \gamma \left(\sigma_{\sum_{i=1}^n \lambda_i X(t_i)} - \sigma_{X(t_0)} \right).
$$

By taking partial derivatives and setting them equal to zero, we get the system of equations

$$
\begin{aligned}
[X(t_j), X(t_0)]_\alpha + \gamma \left[X(t_j), \sum_{i=1}^n \lambda_i X(t_i) \right]_\alpha &= 0, \quad j = 1, \ldots, n, \\
\sigma_{\sum_{i=1}^n \lambda_i X(t_i)} &= \sigma_{X(t_0)}.
\end{aligned}
$$

Suppose that $t_k = t_0$ for some $k \in \{1, \ldots, n\}$. Then (λ^*, γ^*) with

$$\gamma^* = -1 \quad \text{and} \quad \lambda_i^* = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases} \quad i = 1, \ldots, n,$$

is a solution of the system of equations and therefore a critical point of the objective function. Since the MCL predictor is unique, it is given by $\widehat{X(t_0)} = \sum_{i=1}^{n} \lambda_i^* X(t_i)$.

4. Let C be an arbitrary set and $g : C \to \mathbb{R}$ be a function. We set

$$\inf g(C) := \inf\{g(x) : x \in C\},$$
$$M(g, C) := \{x \in C : g(x) = \inf g(C)\}.$$

For a set $M \subset \mathbb{R}^n$ and a sequence $(M_l)_{l \in \mathbb{N}}$ of subsets of $\mathbb{R}^n$, we define

$$M = \lim_{l \to \infty} M_l \iff \lim_{l \to \infty} d(x, M_l) = 0, \ x \in M \text{ and } \liminf_{l \to \infty} d(x, M_l) > 0, \ x \notin M, \tag{A.9}$$

where $d(x, M) := \inf_{m \in M} \|x - m\|_2$, cf. [10, Anhang 1, Bemerkung 1]. Notice that this definition is equivalent to the Hausdorff convergence, see [10, Anhang 1, Satz 4]. The following fact provides sufficient conditions for the stability of the solution of an optimization problem, see [10, Stabilitätssatz 5.4.2].

Let $(S_m)_{m \in \mathbb{N}}$ be a sequence of closed convex subsets of $\mathbb{R}^n$, $n \in \mathbb{N}$, with $\lim_{m \to \infty} S_m := S$. Let $(\Phi_m : \mathbb{R}^n \to \mathbb{R})_{m \in \mathbb{N}}$ be a sequence of convex functions which converges pointwise to some function $\Phi : \mathbb{R}^n \to \mathbb{R}$. If $M(\Phi, S)$ consists of a singleton $\{x\}$, then $x_m \in M(\Phi_m, S_m)$, $m \in \mathbb{N}$, implies $\lim_{m \to \infty} x_m = x$.

Let $(s_m)_{m \in \mathbb{N}} \subset \mathbb{R}^d$ with $\lim_{m \to \infty} s_m = t_0$. We set

$$S_m := \left\{ (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n : \int_E \left| \sum_{i=1}^{n} \lambda_i f_{t_i}(x) \right|^\alpha m(dx) \leq \int_E \left| f_{s_m}(x) \right|^\alpha m(dx) \right\},$$

$$S := \left\{ (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n : \int_E \left| \sum_{i=1}^{n} \lambda_i f_{t_i}(x) \right|^\alpha m(dx) \leq \int_E \left| f_{t_0}(x) \right|^\alpha m(dx) \right\},$$

$$\Phi_m(\lambda) := - \sum_{i=1}^{n} \lambda_i [X(t_i), X(s_m)]_\alpha,$$

$$\Phi(\lambda) := - \sum_{i=1}^{n} \lambda_i [X(t_i), X(t_0)]_\alpha.$$

We notice that we have shown in the second part of the proof that the constraint $\sigma_{\widehat{X(t_0)}} = \sigma_{X(t_0)}$ can be replaced by $\sigma_{\widehat{X(t_0)}} \leq \sigma_{X(t_0)}$. Thus, we can consider S as the set of admissible points of the MCL optimization problem. Furthermore, the previous fact is formulated for minimization problems, but the MCL predictor requires maximization. Therefore, the functions Φ_m and Φ are defined with a negative sign to adjust for this fact. Since the covariation function is continuous, the sequence $(\Phi_m)_{m \in \mathbb{N}}$ converges pointwise to Φ. Also, $M(\Phi, S)$ consists of a singleton $\{\lambda(t_0)\}$ since the MCL predictor is unique. As $\lim_{m \to \infty} s_m = t_0$, there exists some $m_0 \in \mathbb{N}$ such that the set $M(\Phi_m, S_m)$ consists of a singleton for all $m \geq m_0$ since the MCL predictor is unique in a neighborhood of t_0. Clearly, the sets S_m, $m \in \mathbb{N}$, are closed convex subsets of $\mathbb{R}^n$ since the function $\Psi : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\Psi(\lambda) = \int_E \left| \sum_{i=1}^{n} \lambda_i f_{t_i}(x) \right|^\alpha m(dx), \quad \lambda = (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{R}^n,$$

is continuous and convex. It remains to prove that $\lim_{m \to \infty} S_m = S$.
We need to show that $\lim_{m \to \infty} d(\lambda_0, S_m) = 0$ for all $\lambda_0 \in S$ and $\liminf_{m \to \infty} d(\lambda_0, S_m) > 0$ for all $\lambda_0 \notin S$ according to (A.9).

First let $\lambda_0 \in S$. We construct a sequence $(\lambda_m)_{m \in \mathbb{N}}$ with $\lambda_m \in S_m$ such that $\lim_{m \to \infty} \lambda_m = \lambda_0$. Let $\varepsilon > 0$. If $\lambda_0 \in S_m$, then we set $\lambda_m := \lambda_0$. If $\lambda_0 \notin S_m$, then by the definition of S_m, it holds $\Psi(\lambda_0) > 0$. Since Ψ is convex, any local minimum is a global one. Obviously $\min_{\lambda \in \mathbb{R}^n} \Psi(\lambda) = \Psi(0) = 0$ and therefore, λ_0 cannot be a local minimum such that

$$\Psi_{min} := \min_{\lambda \in B_\varepsilon(\lambda_0)} \Psi(\lambda) < \Psi(\lambda_0) < \max_{\lambda \in B_\varepsilon(\lambda_0)} =: \Psi_{max},$$

where $B_\varepsilon(\lambda_0) := \{\lambda \in \mathbb{R}^n : \|\lambda - \lambda_0\|_2 \le \varepsilon\}$. As $\lambda_0 \in S$, we have $\Psi(\lambda_0) \le \kappa(t_0, t_0)$ which implies $\Psi_{min} < \kappa(t_0, t_0)$. Since κ is continuous and $\lim_{m \to \infty} s_m = t_0$, there exists some $m_1 \in \mathbb{N}$ such that $\Psi_{min} < \kappa(s_m, s_m)$ for all $m \ge m_1$. As $\lambda_0 \notin S_m$, we have $\Psi(\lambda_0) > \kappa(s_m, s_m)$ which implies $\kappa(s_m, s_m) < \Psi_{max}$. We have shown that

$$\Psi_{min} < \kappa(s_m, s_m) < \Psi_{max} \quad \forall m \ge m_1.$$

We can find some $\lambda_m \in B_\varepsilon(\lambda_0)$ by the intermediate value theorem such that $\Psi(\lambda_m) = \kappa(s_m, s_m)$ since Ψ is continuous, that is $\|\lambda_m - \lambda_0\|_2 \le \varepsilon$ for all $m \ge m_1$ and $\lim_{m \to \infty} \lambda_m = \lambda_0$.
Now let $\lambda_0 \notin S$. It holds $\Psi(\lambda_0) > \kappa(t_0, t_0)$. Since Ψ is continuous, there exists some $\delta > 0$ such that

$$\Psi(\lambda) > \kappa(t_0, t_0) + \frac{\Psi(\lambda_0) - \kappa(t_0, t_0)}{2}, \quad \forall \lambda \in B_\delta(\lambda_0).$$

Due to the continuity of κ, there exists some $m_2 \in \mathbb{N}$ such that

$$\kappa(s_m, s_m) \le \kappa(t_0, t_0) + \frac{\Psi(\lambda_0) - \kappa(t_0, t_0)}{2}, \quad \forall m \ge m_2,$$

which implies for all $\lambda \in B_\delta(\lambda_0)$ and $m \ge m_2$

$$\Psi(\lambda) > \kappa(t_0, t_0) + \frac{\Psi(\lambda_0) - \kappa(t_0, t_0)}{2} \ge \kappa(s_m, s_m).$$

Therefore $B_\delta(\lambda_0) \cap S_m = \emptyset$ for all $m \ge m_2$ and thus $\liminf_{m \to \infty} d(\lambda_0, S_m) \ge \delta > 0$.

$\square$

References

[1] Brockwell, P. J., Cline, D. B. H., Linear prediction of ARMA processes with infinite variance, Stochastic Process. Appl. 19: 281-296 (1985)

[2] Brockwell, P. J., Mitchell, H., Linear prediction for a class of multivariate stable processes, Stoch. Models 14(1): 297-310 (1998)

[3] Cambanis, S., Soltani, A. R., Prediciton of stable Processes: Spectral and moving average representations, Z. Wahrscheinlichkeitstheorie verw. Gebiete 66: 593-612 (1984)

[4] Cambanis, S., Wu, W., Multiple Regression on Stable Vectors, J. Multivariate Anal. 41: 243-272 (2004)

[5] Chambers, J. M., Mallows, C., Stuck, B.W., A method for simulating stable random variables, J. Amer. Statist. Assoc. 71(354): 340-344 (1976)

[6] DeVore, R. A., Lorentz, G. G., Constructive Approximation, Springer-Verlag, Berlin Heidelberg (1993)

[7] Gallardo, J. R., Matrakis, D., Orozco-Barbosa, L., Prediction of alpha-stable long-range dependent stochastic processes, Adv. Perform. Anal. 3(1): 31-42 (2000)

[8] Hill, J. B., Minimum Dispersion and Unbiasedness: 'Best' Linear Predictors for Stationary ARMA α-Stable Processes, Discussion Papers in Economics, Working Paper No. 00-06, Department of Economics, University at Boulder, Colorado (2000), online available at `http://www.colorado.edu/econ/CEA/papers00/track.pl?p=wp00-6.pdf` (11/30/2010)

[9] Kokoszka, P. S., Prediction of infinite variance fractional ARIMA: Probab. Math. Statist. 16(1): 65-83 (1996)

[10] Kosmol, P., Optimierung und Approximation, Walter de Gruyter, Berlin (1991)

[11] Krige, D. G., A statistical approach to some basic mine valuation problems on the Witwatersrand, J. Chem. Metal. Min. Soc. S. Afr. 52(6): 119-139 (1951)

[12] Lantuéjoul, C., Geostatistical Simulation: Models and Algorithms, Springer-Verlag, Berlin Heidelberg (2002)

[13] Mohammadia M., Mohammadpour A., Best linear prediction for α-stable random processes, Statist. Probab. Lett. 79(21): 2266-2272 (2009)

[14] Molchanov, I., Convex and star-shaped sets associated with multivariate stable distributions, I: Moments and densities, J. Multivariate Anal. 100(10): 2195-2213 (2009)

[15] Moré, J. J., The Levenberg-Marquardt algorithm: Implementation and theory, in Watson, G. (ed.): Numerical Analysis, Lecture Notes in Mathematics 60: 105-116, Springer, Berlin Heidelberg (1978)

[16] Painter, S., Numerical method for conditional simulation of Lévy random fields, Math. Geol. 30(2): 163-179 (1998)

[17] Rosinski, J., Decomposition of stationary α-stable random fields, Ann. Probab. 28(4): 1797-1813 (2000)

[18] Samorodnitsky, G., Taqqu, M. S., Stable Non-Gaussian Random Processes, Chapman & Hall, Boca Raton (1994)

[19] Wackernagel, H., Multivariate Geostatistics, 2nd ed., Springer, Berlin Heidelberg (1998)